



Allen, P., Lourenco, A., & Roberts, L. D. (2016). Detecting duplication in students' research data: A method and illustration. *Ethics and Behavior*, 26(4), 300-311.
<https://doi.org/10.1080/10508422.2015.1019070>

Peer reviewed version

Link to published version (if available):
[10.1080/10508422.2015.1019070](https://doi.org/10.1080/10508422.2015.1019070)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Taylor and Francis at <http://www.tandfonline.com/doi/full/10.1080/10508422.2015.1019070> . Please refer to any applicable terms of use of the publisher

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Detecting Duplication in Students' Research Data: A Method and Illustration

Allen, P. J., Lourenco, A., & Roberts, L. D. (2016). Detecting duplication in students' research data: A method and illustration. *Ethics and Behavior*, 26, 300-311.
doi:10.1080/10508422.2015.1019070

This is an Accepted Manuscript of an article published by Taylor & Francis in *Ethics and Behavior* on 26 Feb 2015, available online:
<http://www.tandfonline.com/10.1080/10508422.2015.1019070>

In compliance with Taylor & Francis's terms and conditions of use, this is the post-print version of the abovementioned work. Readers wishing to cite this paper are encouraged to source the final published version, available at the URL above.

Detecting Duplication in Students' Research Data: A Method and Illustration

Peter J. Allen, Amanda Lourenco, and Lynne D. Roberts

Curtin University, Australia

Author Note

Peter J. Allen, School of Psychology and Speech Pathology, Curtin University,
GPO Box U1987, Perth, Western Australia, 6845. p.allen@curtin.edu.au **

Amanda Lourenco, School of Psychology and Speech Pathology, Curtin
University, GPO Box U1987, Perth, Western Australia, 6845.
amanda.lourenco@student.curtin.edu.au

Lynne D. Roberts, Associate Professor, School of Psychology and Speech
Pathology, Curtin University, GPO Box U1987, Perth, Western Australia, 6845.
lynne.roberts@curtin.edu.au

This research was supported by a grant from the Research and Development
Committee of the School of Psychology and Speech Pathology, Faculty of Health
Sciences, Curtin University.

** Corresponding author.

Abstract

Research integrity is core to the mission of higher education. In undergraduate student samples, self-reported rates of data fabrication have been troublingly high. Despite this, no research has investigated undergraduate data fabrication in a more systematic manner. We applied duplication screening techniques to 18 data sets submitted by psychology honours students for assessment. Although we did not identify any completely duplicated cases, there were numerous partial duplicates. Rather than indicating fabrication however, these partial duplicates are likely a consequence of poor measure selection, insufficient data screening and/or participant characteristics. Implications for the teaching and supervision of honours students are discussed.

Keywords: data fabrication; data duplication; research, ethics

Detecting Duplication in Students' Research Data: A Method and Illustration

Academic integrity is core to the mission of the higher education sector. Research misconduct, which has been widely defined as “fabrication, falsification, or plagiarism in proposing, performing, or reviewing research, or in reporting research results” (Office of Science and Technology Policy, 2000), fundamentally undermines this mission. Of these three forms of research misbehaviour, fabrication (making up data) and falsification (changing data) are often seen as most problematic, as they directly distort scientific knowledge and the decisions based on it (Steneck, 2006). Both have received considerable media attention in recent years (Anderson, Shaw, Steneck, Konkle, & Kamata, 2013).

For example, data fabrication in biomedical research hit the headlines in 2005/6 when it was revealed that South Korean human stem-cell researcher, Woo Suk Hwang, had invented much of the data on which two landmark papers published in *Science* (and later retracted; Hwang et al., 2004, 2005) were based (Cyranoski, 2006). Australian obstetrician William McBride was found guilty of data falsification in research he claimed demonstrated that Debendox, a morning sickness drug, caused birth deformities (Milliken, 1993). Despite evidence suggesting that cases like these occur most commonly in biomedical research (Fanelli, 2009; Grieneisen & Zhang, 2012; Stroebe, Postmes, & Spears, 2012), our own field, psychology, has not escaped scandal. In one recent example, Dutch social psychologist, Diederik Stapel, was exposed as having published over 50 papers based on fraudulent data (Stroebe et al., 2012).

Although it is tempting to conclude that cases like these represent isolated ‘bad apples’ in an otherwise honest system, there is evidence to indicate that the problem is somewhat more widespread. For example, Steneck (2006) calculated that around .001%

of scientists in the US are found guilty of misconduct by a federal oversight agency each year. Based on the ratio of retractions to articles indexed in the PubMed database, Claxton (2005) estimated that between .002% and .02% of published papers report fraudulent data. In a comprehensive review of retraction notices published in 42 scholarly databases between 1980 and 2010, Grieneisen and Zhang (2012) found that retraction rates ranged from .02% to 5.62% across 1,796 unique journal titles, and had increased rapidly in the last 10 years (see also, Steen, 2012). Questionable data or interpretations (including data fabrication and falsification) were cited as the reasons for retraction in 43% of the cases they observed (Grieneisen & Zhang, 2012). Similarly, Fang, Steen, and Casadevall (2012) reported 67.4% of 2047 retractions of articles indexed by PubMed were attributable to misconduct, with 43.4% identified or suspected as fraudulent. Although troubling, these figures almost certainly underestimate the true scope of the problem, considering that rates of confirmed misconduct are a poor proxy for actual rates of misconduct, and cases that are reported and investigated likely only represent 'the tip of the iceberg' (Steneck, 2006).

When self-report data are used to estimate research misconduct rates, the figures are somewhat higher. For example, in a survey funded by The Office of Research Integrity in the US, the Gallup Organization (2008) asked just one randomly selected principal investigator at each of over 4,000 unique schools/departments to report on the research misconduct they had observed in their own workplaces during the previous three years. With a response rate of over 50%, they estimated that around 1.5% of research conducted in the US each year involves some form of serious misconduct. They further estimated that around 60% of that misconduct involves either data fabrication or falsification (Gallup Organization, 2008). In a recent meta-analysis of 18

surveys, Fanelli (2009) found that 1.97% of scientists admitted to having fabricated or falsified data at least once, while nearly 15% reported observing colleagues engaged in this practice. Although these figures almost certainly include some duplicate cases, with the same instances of misconduct being reported by multiple researchers on some occasions (Strobe et al., 2012), Fanelli (2009) argued that they are probably still conservative estimates of the true prevalence of research misconduct within the scientific community. This conclusion is supported by at least three lines of evidence. First, people routinely underreport their own criminal and socially sensitive/undesirable behaviours in self-report surveys (e.g., Farrington, 2001; Krumpal, 2013; Magura & Kang, 1996; Tourangeau & Yan, 2007). This is especially so when social expectations are inconsistent with the behaviours under investigation. This is the case in science, where researchers are expected to act with integrity, a value wholly inconsistent with data fabrication and falsification. Second, researchers report a much higher degree of 'willingness' to engage in future misconduct than they report actual past misconduct (e.g., Eastwood, Derish, Leash, & Ordway, 1996). Third, when researchers are provided with incentives for honesty (e.g., a donation to a charity of their choice, calculated dependent on the estimated truthfulness of their answers), they self-report higher levels of questionable research practices (John, Loewenstein, & Prelec, 2012). John and colleagues (2012) argue that prevalence estimates derived from these methods are likely to be more valid than straight self-reports, a conclusion which is hinged on knowledge that people are less likely to report doing socially sensitive things they haven't done, than not report things they have done (Krumpal, 2013).

Although rates of data fabrication and falsification amongst professional scientists are concerning, they pale into insignificance when compared to self-reported

rates of serious research misconduct amongst university students. For example, some 19% of McCabe's (2005) online sample of over 46,000 North American undergraduate students self-reported fabricating or falsifying laboratory data at least once in the previous year. Eight percent admitted to falsifying research data during the same time period. Amongst his sample of over 7,000 graduate students, the rates of falsifying laboratory and research data were 7% and 4% respectively (McCabe, 2005). A large Australian study by Brimble and Stevenson-Clarke (2005) produced broadly comparable findings, with over 21% of nearly 1,200 (predominantly undergraduate) students indicating that they had falsified the results of their own research at least once. This rate was several times higher than the prevalence estimates provided by academic staff at the same universities. Furthermore, and in contrast to the academic staff, the majority of the students saw falsification as only 'minor cheating' at worst.

In several smaller studies, the numbers are even more troubling. For example, over 60% of Franklyn-Stokes and Newstead's (1995) sample of UK undergraduate science students confessed to inventing or altering research data in the past. Sixty-seven percent of Lawson, Lewis, and Birk's (1999/2000) US sample of biology, chemistry and anatomy undergraduates indicated that they manipulate or make up data at least 'sometimes'. Finally, nearly everyone in Davidson, Cate, Lewis, and Hunter's (2000) US sample of undergraduate biology and chemistry students admitted to manipulating data at least 'often', with many indicating that they do it 'almost always'. The most common reason cited for manipulating data was to obtain a better grade (Davidson et al., 2000).

Although these findings are quite diverse, and varied operational definitions and methodologies make it difficult to pin-down the exact extent of data fabrication and

falsification amongst undergraduate students, when looked at in combination, they nevertheless suggest that these are reasonably common practices throughout large segments of the student population. This is concerning for a number of reasons. For example, cheating “threatens the equity and efficacy of instructional measurement” (Brimble & Stevenson-Clarke, 2005, p. 20), as it can artificially inflate the grades of students who cheat, relative to those who do not. This is especially pertinent in the natural sciences, where producing data that ‘support the hypothesis’ is often directly rewarded with higher grades (Lawson et al., 1999/2000). Furthermore, cheating may result in impoverished learning, leaving students less equipped to deal with more advanced topics in their chosen subject areas (Brimble & Stevenson-Clarke, 2005). Beyond such consequences at the individual and institutional levels, cheating also has broader societal implications. For example, broad awareness of cheating may undermine public trust in entire professions (Marsden, Carroll, & Neill, 2005). For instance, it is not difficult to imagine some people becoming wary of visiting all doctors in the weeks following publication of news stories exposing cheating at a local medical school. Furthermore, dishonesty in college is known to correlate strongly with dishonesty in the workplace (e.g., partial $r > .60$ in Nonis & Swift, 2001), which suggests that dishonest behaviours learned or practiced while studying may transfer or generalise to other contexts. Finally, these findings are all derived from self-report data, which are subject to selection, social desirability and other biases (Crown & Spiller, 1998; Macfarlane, Zhang, & Pun, 2012), and thus may actually underestimate the true scope of academic misconduct in the university student population.

Consequently, a few researchers have attempted to measure student academic misconduct via more objective methods. For example, Pullen, Ortloff, Casey, and Payne

(2000) collected 62 ‘cheat sheets’ they found discarded around their university campus, which they then content analysed. They found that business students tended to be over-represented in their sample, and the ‘typical’ cheat sheet was small enough to easily conceal in the palm of a hand, and densely packed with organised lists of facts. Ward and Beck (1990) asked students to self-score multiple-choice exams in class, and then compared these self-scores to the students’ actual performance on the exam (as determined by Scantron scoring). Some 28% of the sample self-scored themselves higher than their actual scores. By way of contrast, no students gave themselves a score lower than their actual score. Karlins, Michaels, and Podlogar (1988) compared students’ written assignments to those submitted in a previous semester, and found that around three percent had been ‘recycled’ (either by the same, or a different student). Finally, Martin, Rao, and Sloan (2009) reviewed business administration students’ Turnitin similarity reports to detect plagiarism on written assignments, a practice that is now common with Turnitin reporting use of their text matching software in more than 3,500 higher education institutions (iParadigms, 2013). After Martin and colleagues (2009) screened each report to ensure the absence of any false-positives, they found that 61% of their sample met their threshold for plagiarism (at least three percent of the assignment matching a source in the Turnitin database). While these studies provide interesting insights into how students may cheat in some narrowly defined circumstances, they shed little light on the broader scope of the problem. Furthermore, convenience sampling and the esoteric contexts in which they were conducted make generalisation difficult. Finally, we are not aware of any research that has attempted to objectively measure practices suggestive of data fabrication and falsification within a student sample.

There are, however, a number of statistical methods that are available for revealing the possibility of such practices. For example, Evans (2001) describes various univariate, bivariate and multivariate techniques (statistical and graphical) that are useful for screening data derived from clinical trials. Similar methods are discussed by Buyse and colleagues (1999), and illustrated with both fraudulent and legitimate data in Al-Marzouki, Evans, Marshall, and Roberts (2005). Whilst valuable, many of these techniques rely on comparing the characteristics of randomised groups at baseline, which limits their applicability to experimental research for which pre-intervention data are available. When it comes to other types of research data, the literature is less well developed.

Nevertheless, some of the techniques described in Evans (2001) and elsewhere do have broader applicability. One such technique has been developed by Blasius and Thiessen (2012), and can be used to quickly screen data sets for either partially or completely duplicated cases, which may indicate data fabrication. This technique relies on Principal Components Analysis (PCA), and after establishing the rationale for the current research, we will describe its application to the sort of data sets typically produced by researchers (including student researchers) working in the behavioural and social sciences.

The Current Research

In psychology, the honours program is the primary route to both professional (e.g., clinical, organisational, counselling psychology etc.) and research careers, and often provides the first opportunity for students to conceptualise a research project and collect a substantial amount of raw research data. Practices developed during honours may continue into postgraduate studies, and working life (Nonis & Swift, 2001). We are

not aware of any published research on either the self-reported rates of data fabrication or falsification amongst honours students, or the systematic analysis of honours students' raw data for the presence of possible indicators of fabricated or falsified data. The recent development of statistical techniques for detecting certain patterns of data within a broad range of datasets (Blasius & Thiessen, 2012) makes such analysis possible.

The aim of the current research was to systematically examine a sample of psychology honours students' data sets for characteristics which could be attributable to data fabrication. The specific characteristic in question was the presence of either partially or completely duplicated cases. The two research questions driving this research were: (1) is there any evidence of data duplication in the data sets students present with their honours dissertations? (2) If any duplication is found, what are its likely causes?

Method

Participants

In 2012, 32 psychology honours students submitted dissertations in our school, which is situated in a medium-sized, medium-ranked Australian university. Of these, 12 students worked with existing data sets, or collected qualitative data, and were thus excluded from the current sample. A further two students were excluded, as one did not include raw data as a dissertation appendix, and the other's data set was very small, which would have prevented running the analyses described below. Therefore, we analysed a final convenience sample of 18 psychology honours students' dissertations and raw quantitative data sets.

In our school, completed honours dissertations (including raw data sets, which are typically included as digital appendices to each dissertation) are collected in our library, and thus become a part of the public record. Consequently, informed consent was not required from students prior to including their work in our sample (and seeking it would have undermined the intention of the research). Before examining any data sets, the second author (who was not familiar with the 2012 honours students) anonymised them, by giving each a randomly generated file name, and re-naming each variable in each file with a generic code (excluding age and gender). Before examining any dissertations, they were reduced and anonymised by a research assistant. This involved extracting only those sections relevant to data analyses and measurement, and then hashing out any variable names. Together, these processes ensured that we would be unable to link dissertations or data sets back to specific students in the event that any possible fabrication was suspected, and that there could be no adverse consequences to students as a result of their work being included in this research. This research conformed to the guidelines for ethical conduct in human research articulated by the Australian National Health and Medical Research Council (2007), and was approved by our local Human Research Ethics Committee prior to commencing.

Materials and Procedure

A typical honours level data file will contain at least 40 to 50 variables (and often many more), with a variety of response formats (e.g., dichotomous, 5-point Likert, 7-point Likert etc.). The probability of two identical strings of data more than several items long occurring due purely to chance within one of these files is extremely small. For example, if we were to take a 10-item segment of a given data file, and each of those 10-items used a 6-point response format, there are theoretically 10^6 (i.e., 1

million) possible permutations of responses. Consequently, the likelihood of two or more conscientious respondents producing exactly the same permutation of 10 responses within a small honours data file (typically no longer than a couple of hundred cases) due to chance is extremely small. Therefore, it is reasonable to conclude that the presence of such duplicates is probably not due to chance. Furthermore, if there are several sets of duplicates, further investigation to ascertain their likely causes is certainly warranted. These are the basic assumptions on which the following procedure (as first described by Blasius & Thiessen, 2012) was based.

First, we took variables 6 through 15 of each data file (assuming that if deliberate duplication had occurred and the student had attempted to hide their actions, the first and/or last several items of the data file are those most likely to have been modified) and calculated the maximum number of theoretically possible permutations of responses. If the maximum number of permutations possible was below 1 million, successive variables were sampled until this threshold was reached. Next, we subjected the sampled variables to PCA in SPSS (version 20), and saved the component scores for the first component extracted. Note that strings of identical (duplicate) responses will yield identical component scores. (Indeed, this will be the case, regardless of which extraction method is used.) Third, the component scores were graphed, as illustrated in Figure 1. This allowed us to quickly identify whether or not any cases shared component scores. If each case had a unique component score (as illustrated in the left panel of Figure 1), there was no evidence of duplication, and the analysis stopped. However, if any component score was shared by two or more cases (as can be seen in the right panel of Figure 1), they were flagged as ‘potentially problematic cases’ (PPCs), and a similarity index was recorded for each. This similarity index was

calculated by dividing the number of variables the cases were identical on by the total number of variables in the data file, and then multiplying by 100 to derive a percentage. In situations where three or more cases had the same component score, pairs of cases were compared, and only the highest similarity index was recorded against each. The process described above was then repeated from the other end of the data file, to ensure that all PPCs were captured.

<insert Figure 1 about here>

To investigate the possible causes of the identified PPCs in greater detail, the data sets were systematically examined with reference to the dissertation extracts. Specifically, we studied the patterns of responses across each measure in each data set, and documented the nature of each instance where PPCs clustered. Each instance was then cross-referenced against the relevant dissertation to determine whether or not the student author had identified the issue and, if so, the reasons they attributed it to.

Results

The percentage of PPCs in each of the 18 analysed student data files ranged from 0% through to 46.92%, as illustrated by the bars in Figure 2. Sixteen of the 18 data files contained at least some PPCs, but none contained more than 50%. Furthermore, none of the data sets contained fully duplicated cases (i.e., cases with a 100% similarity index), although three contained cases with similarity indices in excess of 90% (as illustrated by the solid line in Figure 2), and 15 contained cases with similarity indices above 50% (as illustrated by the dotted line).

<insert Figure 2 about here>

When the data sets were individually examined with reference to their corresponding dissertations, most PPCs were characterised by one of two features: (1)

long runs of responses to items within a multiple-item measure at either the ceiling or floor; or (2) patterned responses to items on one or more multiple-item measures. As evidence for (1), the majority of items on six different measures had medians at either the lowest or highest response option (range = 50% to 77% of items). A seventh measure had extreme medians on 22% of items. All seven of these measures contained at least one item with more than 70% of all responses at either the ceiling or floor. Regarding (2), eight cases of patterned responding (e.g., 2-2-2-2 or 2-4-2-4) were identified. In the worst instance, a full 10.6% of respondents to one 8-item measure provided an identical response to every item. Only two students identified any problems associated with floor/ceiling effects or patterned responding in their dissertations. The first indicated that one case was removed prior to analysis due to patterned responding, while the second reported on the extreme skew of some item level data.

Discussion

Using the methods described by Blasius and Thiessen (2012), we examined 18 data files submitted with dissertations for assessment by psychology honours students for the existence of partially or completely duplicated cases. Although we did not identify any completely duplicated cases, there were many partial duplicates. Partial duplicates are pairs or sets of cases with matching strings of data that are too long to be reasonably attributable to chance. There are a number of possible explanations for these findings, the majority of which do not suggest any nefarious intent on the part of the students responsible for collecting the data.

First, it may suggest poor measure selection by student researchers. Closer examination of each data file revealed several which were plagued by stereotypical response sets (e.g., responding with a series of 1s on a multiple-item Likert scale). Such

response sets could result from using a measure that is unsuitable for the target population (e.g., a measure designed to assess the severity of schizophrenic symptoms being used with a non-clinical population), or a measure containing items that were difficult to comprehend or irrelevant for sections of the target population.

Second, it may suggest participant fatigue, carelessness and/or socially desirable responding, particularly in online survey-based studies administered through a human subject pool, where participants were identifiable (to permit allocation of course credit) and sometimes asked in excess of 130 questions. Research indicates that this combination of factors may be particularly conducive to careless responding (Meade & Craig, 2012), and that identified surveys tend to elicit higher levels of socially desirable responding than anonymous surveys (Dodou & de Winter, 2014). No students identified these issues as potential concerns within their dissertations.

Third, it may suggest insufficient data screening by student researchers, who ought to have removed cases that had obviously not responded conscientiously (e.g., participants with very short completion times who selected the middle response option for virtually every question) prior to running hypothesis tests. Failure to do so introduces unnecessary error variance into data, which reduces statistical power and the likelihood of detecting meaningfully sized effects (Maniaci & Rogge, 2014). It should be noted that such carelessness and/or incompetence, whilst not representing research misconduct, may still be considered unethical (Wasserman, 2013).

Finally, the partially duplicated cases that we observed could also suggest fabrication coupled with a small amount of data point adjustment, whereby one or two data points were modified post-duplication to ensure that the fabricated cases were not completely identical. All three data sets with similarity indices above 90% contained

cases that differed only on age and/or gender, and one other variable unique to the data set in question.

It should be noted that the explanations presented above are somewhat speculative. It is impossible to fully establish their veracity without interviewing or surveying the researchers and participants involved in each study. It is also important to note that we only considered one specific type of data fabrication (duplication by means of cutting-and-pasting), and so cannot discount the possibility that the student researchers in our sample engaged in other, subtler, fraudulent behaviours. Methods to detect these fraudulent behaviours are documented elsewhere (e.g., Buyse et al., 1999; Evans, 2001). Furthermore, the absence of any egregious data duplication in this small ($N = 18$) convenience sample of Australian psychology honours students should not be used to conclude that undergraduate researchers elsewhere do not engage in such practices. Indeed, several self-report studies suggest that such behaviour may be more common in the natural and physical sciences, where there is a greater expectation that hypotheses will be supported, particularly in straight replications of prior studies (Davidson et al., 2000; Franklyn-Stokes & Newstead, 1995; Lawson et al., 1999/2000). This is an obvious avenue for future research, which should aim to investigate the prevalence of data duplication in larger samples from a variety of disciplines and institutions. Such research should also consider additional plausible causes and correlates of data duplication, including supervisor experience, student ability and various contextual factors. Indeed, one such factor that may have reduced the tendency of this particular cohort of psychology honours students engage in questionable research practices, relative to those from previous of successive years, was the salience of the Diederik Stapel case at the time (see Stroebe et al., 2012). In fact, an article authored by

Stapel was the focus of a major assignment in an advanced research methods course, which many members of the sample completed concurrently to their dissertation research. The article was retracted by the publisher shortly before the assignment submission date, resulting in extended class discussion on academic integrity and research misconduct.

Furthermore, future research should consider methods of assessing the reliability, sensitivity and validity of the techniques described in this paper, which we have not sufficiently addressed. For example, if the method is valid it should be able to detect duplicate cases randomly inserted into a random subset of existing data files by a second, independent researcher. Alternatively, if duplicate cases are an indicator of misconduct, and if students more likely to engage in misconduct are also more likely to engage in other ethically questionable academic behaviours (e.g., Broeckelman-Post, 2009), then dissertations have been identified as problematic for other reasons (e.g., plagiarism) should have a higher probability of also containing duplicate cases than those which have been passed without concern.

Despite the above limitations, there are a number of pedagogic implications that emerge from this research, relating particularly to the teaching and supervision of honours students across disciplines. First, if poor measure selection is responsible for a lack of variability in student data, efforts can be made in class to stress the importance of selecting measures that are valid for the population with which they will be used, and pre-testing them thoroughly in times of doubt. Similar efforts can be made to emphasise the importance of brevity when selecting measures for survey research, particularly if the intention is to administer them online (either anonymously or through subject pool management software), in order to reduce the likelihood of respondent fatigue and

careless responding. Furthermore, time can be spent with students discussing the conceptual importance of data screening and cleaning, as well as the actual mechanics of performing these tasks. Additionally, a number of studies also recommend explicitly addressing research ethics through policies and mentoring (Fisher, Fried, & Feldman, 2009; Fisher, Fried, Goodman, & Germano, 2009). Reminding students about research misconduct has been found to significantly reduce cheating (e.g., McCabe, Trevino, & Butterfield, 2001). Finally, talking with students about the importance of non-significant findings (Ionnadis, 2005; Schooler, 2011) may also help to relieve some of the pressure that many feel to 'reject the null hypothesis'.

In conclusion, this paper describes and illustrates a technique that can be quickly applied to a wide variety of data sets to detect the presence of partially or completely duplicated cases. When it was applied to 18 data sets submitted by psychology honours students for assessment, no complete duplicates were identified, although there were numerous partial duplicates. These duplicates may indicate data fabrication, although in the current context, a more benign etiology appears likely.

References

- Al-Marzouki, S., Evans, S., Marshall, T., & Roberts, I. (2005). Are these data real? Statistical methods for the detection of data fabrication in clinical trials. *BMJ*, 331, 267-270. doi:10.1136/bmj.331.7511.267
- Anderson, M. S., Shaw, M. A., Steneck, N. H., Konkle, E., & Kamata, T. (2013). Research integrity and misconduct in the academic profession. In M. B. Paulsen (Ed.), *Higher education: Handbook of theory and research* (Vol. 28, pp. 217-261). Amsterdam, Netherlands: Springer. doi:10.1007/978-94-007-5836-0_5
- Blasius, J., & Thiessen, V. (2012). *Assessing the quality of research data*. London, UK: Sage.
- Brimble, M., & Stevenson-Clarke, P. (2005). Perceptions of the prevalence and seriousness of academic dishonesty in Australian universities. *The Australian Educational Researcher*, 32(3), 19-44. doi:10.1007/BF03216825
- Broeckelman-Post, M. A. (2009). *Building a culture of academic integrity: The role of communication in creating and changing understandings and enactments of academic integrity* (Doctoral dissertation). Retrieved from http://rave.ohiolink.edu/etdc/view?acc_num=ohiou1242313551
- Buyse, M., George, S. L., Evans, S., Geller, N. L., Ranstam, J., Scherrer, B., ... Verma, B. L. (1999). The role of biostatistics in the prevention, detection and treatment of fraud in clinical trials. *Statistics in Medicine*, 18, 3435-3451. doi:10.1002/(SICI)1097-0258(19991230)18:24<3435::AID-SIM365>3.0.CO;2-O

- Crown, D. F., & Spiller, M. S. (1998). Learning from the literature on collegiate cheating: A review of empirical research. *Journal of Business Ethics*, 17, 683-700. doi:10.1023/A:1017903001888
- Cyranoski, D. (2006). Verdict: Hwang's human stem cells were all fakes. *Nature*, 439, 122-123. doi:10.1038/439122a
- Davidson, E. W., Cate, H. E., Lewis, C. M., Jr., & Hunter, M. (2000, November 19-21). *Data manipulation in the undergraduate laboratory: What are we teaching?* Paper presented at the First ORI Research Conference of Research Integrity, Bethesda, MD. Retrieved from <http://ori.dhhs.gov/documents/papers/davidson.pdf>.
- Dodou, D., & de Winter, J. C. F. (2014). Social desirability is the same in offline, online and paper surveys: A meta-analysis. *Computers in Human Behavior*, 36, 487-495. doi:10.1016/j.chb.2014.04.005
- Eastwood, S., Derish, P., Leash, E., & Ordway, S. (1996). Ethical issues in biomedical research: Perceptions and practices of postdoctoral research fellows responding to a survey. *Science and Engineering Ethics*, 2, 89-114. doi:10.1007/BF02639320
- Evans, S. (2001). Statistical aspects of the detection of fraud. In S. Lock, F. Wells, & M. J. G. Farthing (Eds.), *Fraud and misconduct in biomedical research* (3rd ed., pp. 186-204). London, UK: BMJ Books.
- Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLOS One*, 4(5), e5738. doi:10.1371/journal.pone.0005738

- Fang, F. C., Steen, R. G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 17028-17033. doi:10.1073/pnas.1212247109
- Farrington, D. P. (2001). *What has been learned from self-reports about criminal careers and the causes of offending?* Home Office Online Report retrieved from http://www.crim.cam.ac.uk/people/academic_research/david_farrington/srdrep.pdf
- Fisher, C. B., Fried, A. L., & Feldman, L. G. (2009). Graduate socialization in the responsible conduct of research: A national survey on the research ethics training experiences of psychology doctoral students. *Ethics & Behavior*, 19, 496-518. doi:10.1080/10508420903275283
- Fisher, C. B., Fried, A. L., Goodman, S. J., & Germano, K. K. (2009). Measures of mentoring, department climate, and graduate student preparedness in the responsible conduct of psychological research. *Ethics & Behavior*, 19, 227-252. doi:10.1080/10508420902886726
- Franklyn-Stokes, A., & Newstead, S. E. (1995). Undergraduate cheating: who does what and why? *Studies in Higher Education*, 20, 159-172. doi:10.1080/03075079512331381673
- Grieneisen, M. L., & Zhang, M. (2012). A comprehensive survey of retracted articles from the scholarly literature. *PLOS One*, 7(10), e44118. doi:10.1371/journal.pone.0044118

- Gallup Organization. (2008). *Final report: Observing and reporting suspected misconduct in biomedical research*. Retrieved from http://ori.hhs.gov/sites/default/files/gallup_finalreport.pdf
- Hwang, W. S., Roh, S. I., Lee, B. C., Kang, S. K., Kwon, D. K., Kim, S., ... Schatten, G. (2005). Patient-specific embryonic stem cells derived from human SCNT blastocysts. *Science*, 308, 1777-1783. doi:10.1126/science.1112286
- Hwang, W. S., Ryu, Y. J., Park, J. H., Park, E. S., Lee, E. G., Koo, J. M., ... Moon, S. Y. (2004). Evidence of a pluripotent human embryonic stem cell line derived from a cloned blastocyst. *Science*, 303, 1669-1674. doi:10.1126/science.1094515
- Ionnadis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), e124. doi: 10.1371/journal.pmed.0020124
- iParadigms, LLC. (2013). *Turnitin: Colleges & university*. Retrieved 10 August, 2013, from http://turnitin.com/en_us/customers/college-and-university
- Karlins, M., Michaels, C., & Podlogar, S. (1988). An empirical investigation of actual cheating in a large sample of undergraduates. *Research in Higher Education*, 29, 359-364. doi:10.1007/BF00992776
- Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality and Quantity*, 47, 2025-2047. doi:10.1007/s11135-011-9640-9
- Lawson, A. E., Lewis, C. M., Jr., & Birk, J. P. (1999/2000). Why do students “cook” data? A case study on the tenacity of misconceptions. *Journal of College Science Teaching*, 29, 191-198.

- Magura, S., & Kang, S. -Y. (1996). Validity of self-reported drug use in high risk populations: A meta-analytical review. *Substance Use & Misuse*, 31, 1131-1153. doi:10.3109/10826089609063969
- Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, 48, 61-83. doi:10.1016/j.jrp.2013.09.008
- Martin, D. E., Rao, A., & Sloan, L. R. (2009). Plagiarism, integrity, and workplace deviance: A criterion study. *Ethics & Behavior*, 19, 36-50. doi:10.1080/10508420802623666
- Marsden, H., Carroll, M., Neill, J. T. (2005). Who cheats at university? A self-report study of dishonest academic behaviours in a sample of Australian university students. *Australian Journal of Psychology*, 57, 1-10. doi:10.1080/00049530412331283426
- Macfarlane, B., Zhang, J., & Pun, A. (2012). Academic integrity: A review of the literature. *Studies in Higher Education*. Advance online publication. doi:10.1080/03075079.2012.709495
- McCabe, D. L. (2005). Cheating among college and university students: A North American perspective. *International Journal of Educational Integrity*, 1(1). Retrieved from <http://www.ojs.unisa.edu.au/index.php/IJEI/article/view/14>
- McCabe, D. L., Trevino, L. K., & Butterfield, K. D. (2001). Cheating in academic institutions: A decade of research. *Ethics & Behavior*, 11, 219-232. doi:10.1207/S15327019EB1103_2
- Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17, 437-455. doi:10.1037/a0028085

- Milliken, R. (1993, February 20). 'Thalidomide doctor' guilty of medical fraud: William McBride, who exposed the danger of one anti-nausea drug, has been disgraced by experiments with another, writes Robert Milliken in Sydney. *The Independent*. Retrieved from <http://www.independent.co.uk/news/world/thalidomide-doctor-guilty-of-medical-fraud-william-mcbride-who-exposed-the-danger-of-one-antinausea-drug-has-been-disgraced-by-experiments-with-another-writes-robert-milliken-in-sydney-1474190.html>
- National Health and Medical Research Council, Australian Research Council, & Australian Vice-Chancellors' Committee. (2007). *National statement on ethical conduct in human research 2007 (Updated May 2013)*. Canberra, Australia: Author. Retrieved from <http://www.nhmrc.gov.au/guidelines/publications/e72>
- Nonis, S., & Swift, C. O. (2001). An examination of the relationship between academic dishonesty and workplace dishonesty: A multicampus investigation. *Journal of Education for Business*, 77, 69-77. doi:10.1080/08832320109599052
- Office of Science and Technology Policy. (2000). *Federal research misconduct policy*. Washington, DC: Author. Retrieved from <http://ori.hhs.gov/federal-research-misconduct-policy>
- Pullen, R., Ortloff, V., Casey, S., & Payne, J. B. (2000). Analysis of academic misconduct using unobtrusive research: A study of discarded cheat sheets. *College Student Journal*, 34, 616-625.
- Schooler, J. (2011). Unpublished results hide the decline effect. *Nature*, 470, 437. doi:10.1038/470437a

- Steen, R. G. (2012). Retractions in the scientific literature: Is the incidence of research fraud increasing? *Journal of Medical Ethics*, 37, 249-253. doi:10.1136/jme.2010.040923
- Steneck, N. H. (2006). Fostering integrity in research: Definitions, current knowledge and future directions. *Science and Engineering Ethics*, 12, 53-74. doi: 10.1007/PL00022268
- Stroebe, W., Postmes, T., & Spears, R. (2012). Scientific misconduct and the myth of self-correction in science. *Perspectives on Psychological Science*, 7, 670-688. doi:10.1177/1745691612460687
- Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133, 859-883. doi:10.1037/0033-2909.133.5.859
- Ward, D. A., & Beck, W. L. (1990). Gender and dishonesty. *The Journal of Social Psychology*, 130, 333-339. doi:10.1080/00224545.1990.9924589
- Wasserman, R. (2013). Ethical issues and guidelines for conducting data analysis in psychological research. *Ethics & Behavior*, 23, 3-15.

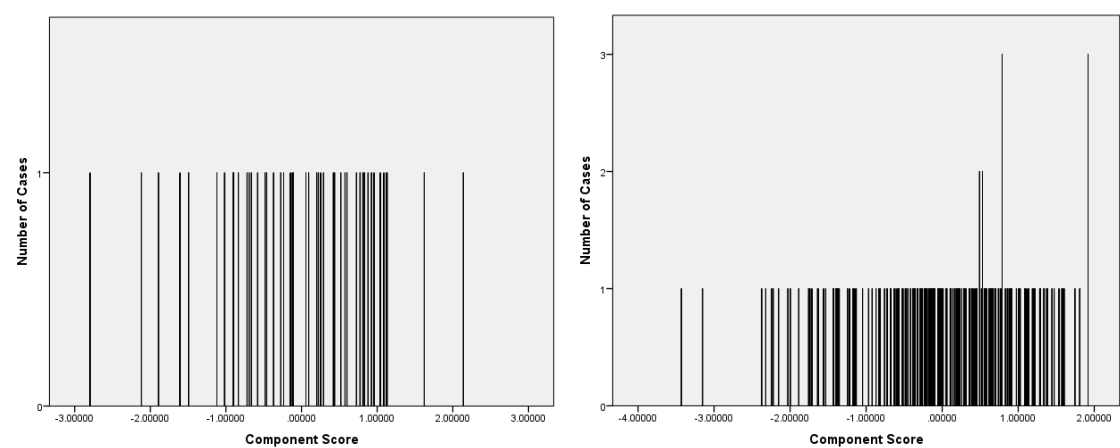


Figure 1. Bar graphs illustrating the absence (left) and presence (right) of cases with identical component scores. Cases with identical component scores have an identical string of responses to the set of variables analysed. There are 10 PPCs visible in the right panel.

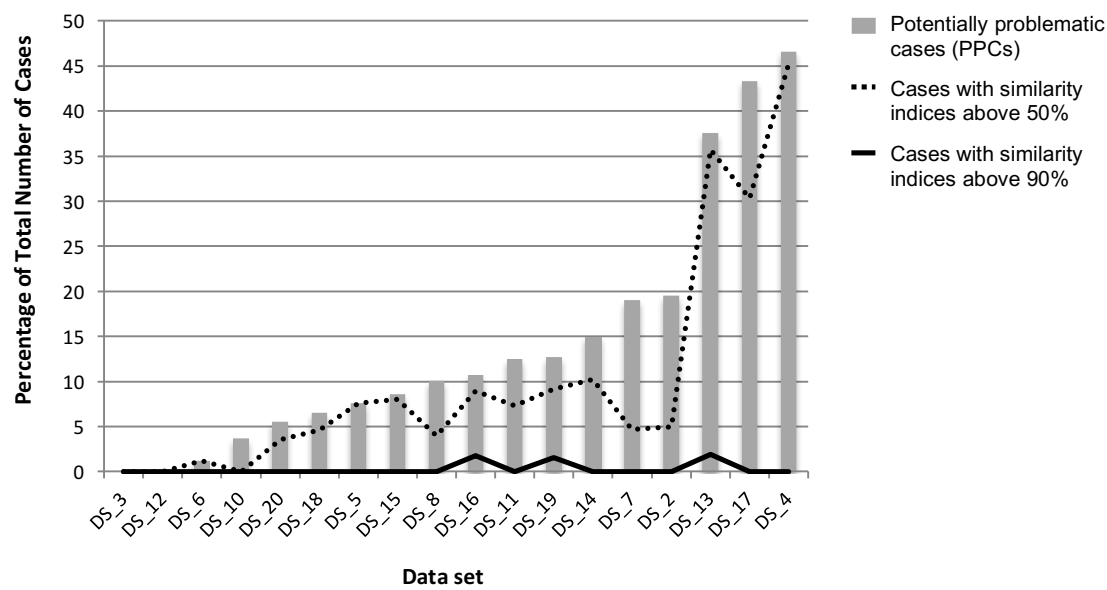


Figure 2. Bar graph illustrating the percentage of PPCs in each data set, along with the percentage of cases in each set with similarity indices above 50% and 90%. Percentages reported relate to the full data sets, as provided by the student researchers. In eight instances, the student researchers reported analyses derived from a reduced data set. However, the only reason cited for excluding cases from analyses was missing data, and component scores were not computed for cases with missing data. Therefore, in all instances, re-running the analyses on the reduced data sets would have inflated the figures reported above.